

Intention-Oriented Process Model Discovery from Incident Management Event Logs

Ashish Sureka
Software Analytics Research Lab (SARL), India
www.software-analytics.in
ashish@iiitd.ac.in

ABSTRACT

Intention-oriented process mining is based on the belief that the fundamental nature of processes is mostly intentional (unlike activity-oriented process) and aims at discovering strategy and intentional process models from event-logs recorded during the process enactment. In this paper, we present an application of intention-oriented process mining for the domain of incident management of an Information Technology Infrastructure Library (ITIL) process. We apply the Map Miner Method (MMM) on a large real-world dataset for discovering hidden and unobservable user behavior, strategies and intentions. We first discover user strategies from the given activity sequence data by applying Hidden Markov Model (HMM) based unsupervised learning technique. We then process the emission and transition matrices of the discovered HMM to generate a coarse-grained Map Process Model. We present the first application or study of the new and emerging field of Intention-oriented process mining on an incident management event-log dataset and discuss its applicability, effectiveness and challenges.

Keywords

Business Process Intelligence (BPI), Hidden Markov Models (HMM), Incident Management Event Logs, Intention Mining, Process Mining, Unsupervised Learning

1. RESEARCH MOTIVATION AND AIM

Process mining consists of extracting or discovering runtime process models from event logs, measuring the extent of compliance between the design time and actual process models, improving process models and analyzing it from multiple perspectives (such as control-flow and organizational perspectives) [4][5]. Intention-oriented process mining is a new and emerging area which is based on the belief that the fundamental nature of processes is mostly intentional (unlike activity-oriented process mining which specifies behaviors in-terms of sequences of tasks and branches) [1][2][3]. Intention-oriented process mining is a relatively unexplored

area except the recent (year 2013 and 2014) and pioneering work on the topic by Khodabandelou et al. [1][2][3]. Discovery of intentional process models from event logs has been proposed for the first time by Khodabandelou et al. and we believe that several more studies are required to throw more light on the applicability and effectiveness of intention-oriented process modeling for solving practical problems encountered by process owners in enterprises. Khodabandelou et al. demonstrate the application of intention-oriented process modeling on two datasets: event logs of developers who committed their activities to Usage Data Collector (UDC) of Eclipse and a laboratory context where an experiment was conducted with University Master students for Entity-Relationship diagrams creation [1][2][3]. The study presented in this paper is motivated by the need to extend the state-of-the-art in the emerging area of intention-oriented process modeling by applying the approach on a large real-world dataset and sharing the findings with the research community. The specific research aims of the work presented in this paper is the following:

1. To investigate the application of intention-oriented process modeling approach on real-world event-log data extracted from incident management systems of an enterprise. To the best of our knowledge, this is the *first study* on the application of intention mining and intention-oriented process model discovery on a large real-world incident management dataset.
2. To examine customization or extension of the approach proposed by the pioneers of intention-oriented process modeling for the given context and application scenario.

2. EXPERIMENTAL DATASET

We conduct our study on large real-world publicly available dataset so that our experiments can be replicated and the results can be used for comparison or benchmarking purposes. The work presented in this paper holds the required *replication standards* ensuring sufficient information for any third party to replicate the results without any additional information from us. We conduct experiments on the publicly available dataset provided by the tenth¹ International Workshop on Business Process Intelligence (BPI). Data collection is one of the most important stage in conducting qualitative research and the quality of result obtained depends both on research design and data gathered. The data

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

¹ <http://www.win.tue.nl/bpi/2014/challenge>

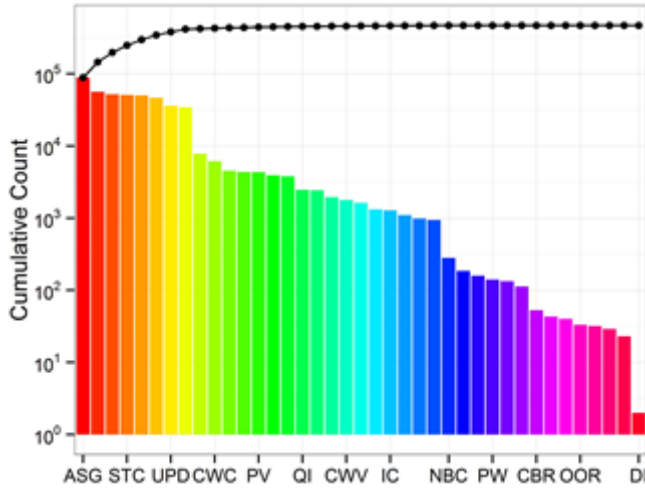


Figure 1: Pareto chart showing the distribution of activities and their cumulative count. Y-axis is in logarithmic scale.

Table 1: Actor, Activity and Timestamp for one of the Cases in the Dataset

DateStamp	Activity	Group
7/1/2013 8:17	Reassignment	01
4/11/2013 13:41	Reassignment	02
4/11/2013 13:41	Update from cust	02
4/11/2013 12:09	Operator Update	03
4/11/2013 12:09	Assignment	03
4/11/2013 13:41	Assignment	02
4/11/2013 13:51	Closed	03
4/11/2013 13:51	Caused By CI	03
4/11/2013 12:09	Reassignment	03
25/09/2013 08:27	Operator Update	03

provided on the BPI workshop website is of high quality as it is peer-reviewed and prepared by experts on the given topic. As an academic, we believe and encourage academic code or software sharing in the interest of improving *openness and research reproducibility*. We release our code and dataset in public domain so that other researchers can validate our scientific claims and use our tool for comparison or benchmarking purposes (and also reusability and extension). Our code and is hosted on GitHub²[Currently not mentioned due to blind review policy] which is a popular web-based hosting service for software development projects. We select GPL license (restrictive license) so that our code can never be closed-sourced.

3. EMPIRICAL ANALYSIS

3.1 Discovering User Strategies from Activities

The Rabobank Group ICT Incident Dataset consists of 46616 incidents or cases and 466737 events. The fields in the event-log dataset are: Incident ID, TimeS-tamp, Incident Activity Number, Incident Activity Type, Assignment

²Currently not mentioned due to blind review policy

Group and KM number (a number related to knowledge document). Table 1 shows the Actor, Activity and Timestamp for one of the Cases in the Dataset. The even-log data in Table 1 shows that several activities are performed by various actors during the workflow and process enactment. Our objective is to apply the Map Miner Method (MMM) that aims at discovering intentions and strategies from event or activity logs and thereby build the actual intention-oriented process model [1][2][3]. The intention-oriented process model is a tool that complements the activity-oriented process model helping the user to better understand the deep nature of the business processes from multiple perspectives. The first step in the process is to model users' strategies in terms of observed activities (present in event logs) using Hidden Markov Models (HMMs). The input to the MMM is the temporal set of user's activities (an activity is an interaction of the user with the information system) during a time-slice wherein the time-slice for a temporal-set or sequence of activity is a specific process trace or case. Figure 1 displays a Pareto chart showing the total number of activities and the distribution of 39 different kind of activities in the dataset (The Y-axis of Figure 1 is in logarithmic scale). The activity distribution is skewed as activities such as Assignment (ASG), Operator Update (OU), Reassignment (RASG), Closed (CLD) and Status Change (SC) have 88502, 56292, 51961, 50145 and 50914 entries respectively whereas activities such as Referred (REF), Problem Closure (PC), OO Response (OOR), Dial-In (DI) and Contact Change (CC) have 29, 40, 33, 2 and 32 entries respectively. Hidden Markov Models (HMMs) are stochastic finite automaton and statistical Markov Models in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states and can be presented as the simplest Dynamic Bayesian Network³. We use Jahmm⁴ library which is a Java implementation of Hidden Markov Model (HMM) related algorithms. Jahmm provides implementation of the Viterbi, Forward-Backward, Baum-Welch and K-Means algorithms in addition to other algorithms related to HMM. We transform the dataset in the format required by the Jahmm library. The source code for data transformation and analysis is available on the GitHub website for the project⁵[Currently not mentioned due to blind-review policy].

We first learn an initial HMM using K-Means learning algorithm. The K-Means learner is initialized with 12 states (number of strategies is equal to the number of states). We specified the number of states as 12 (a heuristic, number of states equal to $\frac{1}{3}$ of the number of different kind of activities) as BWA requires the cardinality of the strategy set \mathbb{S} . We then pass the initial HMM to the Baum Welch Learner which is the most commonly used algorithm to learn the parameters of a HMM (unsupervised learning). The model parameters to be learnt are $\mathbb{H} = \{\mathbb{E}, \mathbb{T}\}$. \mathbb{T} is a list of lists or a square $N \times N$ matrix, whose (i, j) entry gives the probability of transitioning from state i to state j . \mathbb{E} a list of N lists or a matrix with N rows, such that $\mathbb{E}[i, k]$ gives the probability of emitting symbol k while in state i . The behavior of the HMM is captured in \mathbb{H} . We consider a uniform distribution for the probabilities of starting in each initial state. A set of activities that is realized to fulfill a given intention is

³http://en.wikipedia.org/wiki/Hidden_Markov_model

⁴<http://code.google.com/p/jahmm/>

⁵Currently not mentioned due to blind-review policy

Table 2: Table showing the mapping between the underlying 12 user strategies and activities derived from user’s traces recorded during the process enactment

S	π	Activities	Distribution
S ₁	0.07	Reassignment, Communication with vendor	[0.97, 0.03]
S ₂	0.19	Update from customer, Notify By Change, Open	[0.08, 0.01, 0.92]
S ₃	0.09	External Vendor Assignment, Operator Update, Urgency Change, Communication with customer	[0.06, 0.83, 0.02, 0.09]
S ₄	0.16	Assignment	[1]
S ₅	0.17	Closed, Resolved, Quality Indicator Set	[0.93, 0.03, 0.04]
S ₆	0.11	Caused By CI, Reopen	[0.93, 0.07]
S ₇	0.07	Impact Change, Quality Indicator Fixed, Update	[0.03, 0.17, 0.80]
S ₈	0.10	Status Change, Mail to Customer	[0.93, 0.07]
S ₉	0.01	External update, Pending vendor, Problem Closure, Callback Request	[0.20, 0.78, 0.01, 0.01]
S ₁₀	-	Analysis-Research, Description Update	[0.18, 0.82]
S ₁₁	-	Vendor Reference Change, Quality Indicator, Vendor Reference, Incident reproduction	[0.04, 0.69, 0.26, 0.01]
S ₁₂	-	alert stage 1, Service Change	[0.42, 0.58]

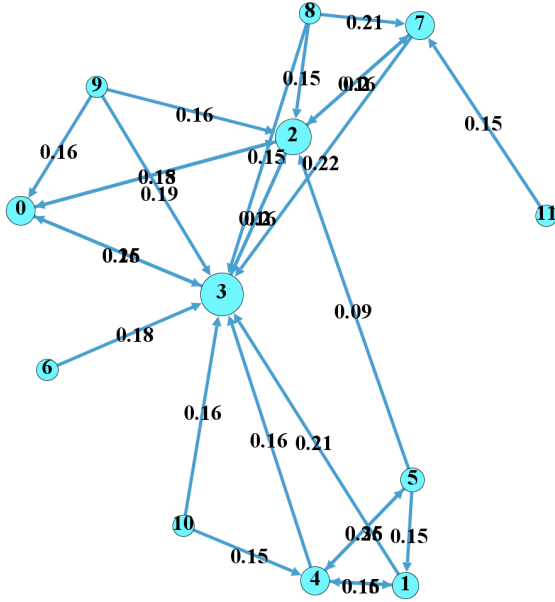


Figure 2: Discovered HMM consisting of 12 states

called as a strategy and thus a strategy comprises of several activities.

Table 2 reveals the discovered mapping between the underlying 12 user strategies and activities derived from user’s traces recorded during the process enactment. As shown in the Table 2, activities Update from customer, Notify By Change and Open constitutes the strategy S₂. Similarly, activity Assignment constitutes the strategy S₄. Table 2 shows the emission probabilities for activities within a strategy. The distribution column shows how many times a given activity appears in a given strategy. The related probabilities are called emission probabilities. For example, The hidden state S₆ generates the activities Caused by CI and Reopen with a probability distribution of 0.93 and 0.07 respectively. The user strategies are hidden and not directly observable (unlike activities). Figure 2 shows the topology of the dis-

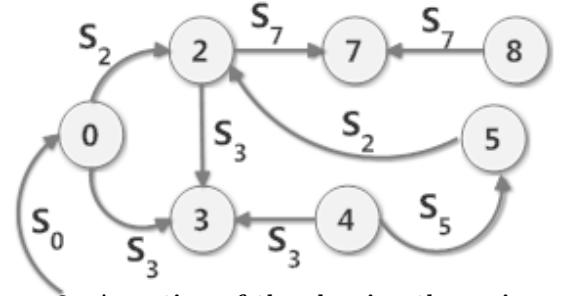


Figure 3: A section of the showing the assignment of strategies to sub-intentions

covered HMM and also some of the HMM parameters. We generated several HMMs models with different numbers of strategies but presenting the HMM with 12 states only as a proof-of-concept due to limited space in the paper. As prescribed in the MMM, we eliminate the elements of the transition matrix T which are smaller than a specified threshold $\epsilon=0.15$. The HMM in Figure 2 consists of 12 nodes where the size of the node is proportional to the number of incoming links. The labels in the edge (value ≥ 0.15) represents the transition probabilities. The thickness of the edge is proportional to the value of the transition probability.

3.2 Discovering Intention-Oriented Process Model

Once the HMM and transition matrix T is derived, the next step consists of assigning each strategy in the matrix a target sub-intention. Each strategy leads to an intention according to the Map formalism. Next step of the method after discovering the HMM is to construct the relationship between strategies according to the transition matrix. Figure 3 shows a fragment of the Map generated as the result of connecting the strategies according to the procedure defined by the Map Miner Method. Strategy S₂ is connected to strategy S₃ and strategy S₇ as there is an element in the transition matrix from S₂ to S₃ and S₂ to S₇.

The next step consists of determining the Start and Stop intentions. The sub-intention(s) for which there is no incoming transition serves as the beginning of the process and corresponds to the Start intention. Similarly, the intention for

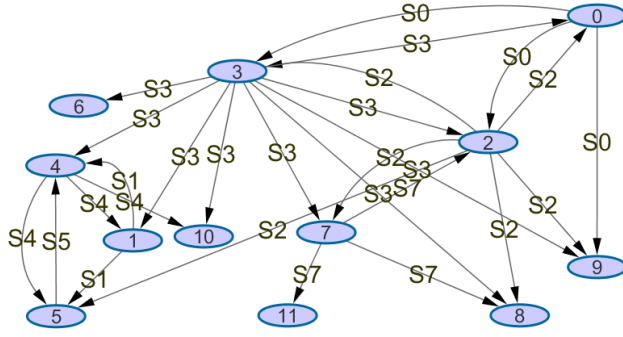


Figure 4: Complete Pseudo-Map consisting of strategies connected to sub-intentions

Table 3: Result of clustering sub-intentions to intentions. CC: Clustering Coefficient, CL: Closeness Centrality, EC: Eccentricity, NC: Neighborhood Connectivity

Node	Cluster	CC	CL	EC	NC
11	C1,C2	0.0	0.5	3	4.0
6	C1	0.0	0.5	3	9.0
3	C1, C2, C3	0.15	0.75	2	3.22
2	C1, C2, C3	0.23	1.0	1	4.16
4	C1, C2	0.25	0.6	3	4.25
5	C1, C2	0.33	0.66	2	4.33
7	C1, C2	0.33	0.75	2	4.75
1	C1, C2	0.5	0.54	3	5.33
10	C1	0.5	0.53	3	6.5
0	C1, C3	0.66	0.75	2	6.0
8	C1, C2	0.83	0.80	2	6.33
9	C1, C3	1.0	0.80	2	6.0

which there is no outgoing transition corresponds to the end of the process and serves as the Stop intention. In our example, we have four nodes 6, 8, 9 and 10 for which there are no incoming transitions. We have node 11 for which there is only one outgoing transition with transition probability of 0.15. We experiment with node 6 as the Start intention and node 11 as the Stop intention. Figure 4 shows the complete pseudo-map consisting of all the strategies connected to sub-intentions. The pseudo-map shows how a sub-intention can be reached from other sub-intentions by following different paths or sequence of strategies. The model containing sub-intentions is called as pseudo-Map and are clustered in-order to group them into clusters of intentions. The parameter N (number of clusters) determines the level of granularity and the number of intentions. After clustering a final Map process model is rebuilt. We use maximal clique-based EAGLE algorithm implemented in Cytoscape⁶ to identify the clusters and group sub-intentions to intentions. Applying EAGLE algorithm is our context specific customization to the MMM. We use a clique-size threshold of 3 and a complex-size threshold of 2. Table 3 shows the result of clustering sub-intentions to intentions. We discover 3 intentions and the Table 3 shows the nodes and strategies belonging to each

⁶ <http://www.cytoscape.org/>

of the intentions.

4. THREATS TO VALIDITY

The quality of discovered process map depends on the number of states defined in the HMM based on a heuristic and experience of an expert. The number of BWA iterations can also influence the outcome and in our case-study we set 50 iterations which may not be optimal. Similarly, clustering sub-intentions to intentions requires parameters setting which may not be optimal for our case-study.

5. CONCLUSIONS

Discovering strategies and intentions (intention-oriented process maps) allows understanding the service desk and IT operation behavior during the incident resolution process. The process model discovers multiple strategies and intentions and Table 1, Figure 2, 3 and 4 shows that service desk have selected different sequence of strategies or paths with different probabilities to fulfill their objectives or goals. We observe that activities External Vendor Assignment, Operator Update, Urgency Change, Communication with customer combined is a prevent strategy followed by several service-desk operators. Activities Reassignment, Communication with vendor, Assignment are part of one intention or cluster. A process owner can analyze the behavior of operators using the intention-oriented map in order to understand how, why and with which probabilities service desk operators address the reported incidents. The map also shows which paths are more or less taken and where are system bottlenecks. For example, network analysis reveals that the network diameter is 3, network density is 0.318, network centralization is 0.6 and characteristic path length is 1.803. Understanding best practices and best path to fulfill an intention is an important aspect of intention-oriented process mining. The discovered map connecting start state to end state and various sub-intentions and intentions can be used to provide recommendations to the operators in-terms of the best path to achieve a certain goal.

6. REFERENCES

- [1] G. Khodabandelou, C. Hug, R. Deneckère, and C. Salinesi. Unsupervised discovery of intentional process models from event logs. In *MSR*, MSR 2014, pages 282–291, New York, NY, USA, 2014. ACM.
- [2] G. Khodabandelou, C. Hug, R. Deneckère, and C. Salinesi. Process mining versus intention mining. In *Enterprise, Business-Process and Information Systems Modeling*, volume 147 of *LNBIS*, pages 466–480. Springer Berlin Heidelberg, 2013.
- [3] G. Khodabandelou, C. Hug, and C. Salinesi. A novel approach to process mining: Intentional process models discovery. In *Research Challenges in Information Science (RCIS)*, pages 1–12, May 2014.
- [4] W. van der Aalst, T. Weijters, and L. Maruster. Workflow mining: Discovering process models from event logs. *IEEE Trans. on Knowl. and Data Eng.*, 16(9):1128–1142, Sept. 2004.
- [5] W. M. P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer Publishing Company, Incorporated, 1st edition, 2011.